Ecol 145 Assignment 10
Dahl Winters
4/14/06

Question 1

In Wednesday's class we found it necessary when adding the variable **climate.3** to center the variable **lntemp** in order to get the optimization algorithm to converge. To simplify interpretation we then used as our level-1 predictor, **lntemp**–log(15).

1. Fit a multilevel model with random slopes and intercepts in which the response is **lnPLD** and the only predictor is the level-1 predictor **lntemp**. This is model2 from class.

```
sp<-
read.table('http://www.unc.edu/courses/2006spring/ecol/145/001/data/lab11/74s
pecies.csv', header=TRUE, sep=',')

model2<-lme(lnPLD~lntemp,random=~lntemp|species,data=sp,method='ML')
```

2. Next fit a model that is identical in every respects to this model except that the level-1 predictor is **lntemp**-log(15), a centered version of log temperature. This is model2.5 from class.

```
model2.5<-lme(lnPLD~I(lntemp-log(15)), random=~I(lntemp-log(15))|species,
data=sp,method='ML')
```

3. Show with minimal algebraic manipulation that the fixed effect estimates you get back (both the intercept and slope) are numerically exactly the same in the two models. This proves that centering is just a numerical trick to achieve better numerical stability. It does not actually change the model being fit.

```
summary(model2)
Fixed effects: lnPLD ~ lntemp
                Value  Std.Error  DF    t-value p-value
(Intercept)  7.088666 0.29554213 143   23.98530       0
lntemp      -1.454417 0.08465553 143 -17.18041        0

summary(model2.5)
Fixed effects: lnPLD ~ I(lntemp - log(15))
                         Value Std.Error  DF    t-value p-value
(Intercept)           3.150033 0.1107243 143   28.44933       0
I(lntemp - log(15)) -1.454417 0.0846555 143  -17.18041        0
```

Substituting the fixed effect values β_0 and β_1 from both models

```
Model2     y = 7.088666 - 1.454417x


Model2.5   y = 3.150033 - 1.454417(x-log(15))
             = 3.150033 - 1.454417x + 1.454417*log(15)
             = 3.150033 - 1.454417x + 3.938634
             = (3.150033 + 3.938634) - 1.454417x
             = 7.088667 - 1.454417x
```

Question 2

Assess whether the variable **feeding.type** affects the linear relationship between lnPLD and lntemp. Decide if its primary effect is on the intercept, slope, both the intercept and slope, or neither the intercept or slope.

If feeding type affects just the intercept, then it appears in the first level equation as an additional predictor.
```
model3<-lme(lnPLD~I(lntemp-log(15))+feeding.type,random=~I(lntemp-
log(15))|species,data=sp,method='ML')
```

If feeding type affects just the slope, then it appears in the first level equation as an interaction term with the slope (x:z).
```
model3a<-lme(lnPLD~I(lntemp-log(15))+I(lntemp-log(15)):feeding.type,
random=~I(lntemp-log(15))|species,data=sp,method='ML')
```

If feeding type affects both the intercept and the slope, then it appears twice in the level 1 equation – once for the intercept term (z) and once for the slope term (x:z), so that the level 1 equation is Y~x+x:z.
```
model3b<-lme(lnPLD~I(lntemp-log(15))*feeding.type,random=~I(lntemp-
log(15))|species,data=sp,method='ML') # * means x, z, and x:z.
```

If feeding type has no effect on the intercept or slope, then this is model 2.5 from Question 1:
```
model2.5<-lme(lnPLD~I(lntemp-log(15)), random=~I(lntemp-log(15))|species,
data=sp,method='ML')
```

Finding the best model using AIC
```
sapply(list(model3,model3a,model3b,model2.5),AIC)
[1] 203.2484 211.8966 204.7071 210.4372
```

The AIC for model3 (the model with feeding type affecting intercept only) has the lowest AIC, so the primary effect of feeding.type is to modify the intercepts of the lines (lnPLD vs. centered lntemp) of the two feeding types.

Checking whether feeding.type is a significant predictor
The AIC suggests model3 is the best model, but we should still check to see whether feeding.type is a significant predictor. The Wald significance test performed for feeding.typeP is a test of whether feeding.type (β_2) is different from 0. Since the p-value is significant (0.0027 < 0.05), feeding.type should be a predictor and thus the lines for each species should have different intercepts.

```
summary(model3)
Fixed effects: lnPLD ~ I(lntemp - log(15)) + feeding.type
                          Value   Std.Error   DF    t-value  p-value
(Intercept)            2.6374882 0.19500883  143   13.524968  0.0000
I(lntemp - log(15))   -1.4489601 0.08480383  143  -17.086022  0.0000
feeding.typeP          0.6773869 0.21762860   72    3.112582  0.0027
```

Question 3

For the model you chose to be best in Question 2, do the following.

1. Write the model as a two-stage model, i.e., write down the level-1 equation and the level-2 equations.

Level 1: $Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \varepsilon_{ij}$
Level 2: $\beta_{0i} = \beta_0 + u_{0i} + \beta_2 z_i$
  $\beta_{1i} = \beta_1 + u_{1i}$

2. Write the model as a composite model, i.e., write down the model as a single equation in which the fixed effect and random effects are segregated into different portions of the equation.

Composite: $Y_{ij} = (\beta_0 + \beta_1 x_{ij} + \beta_2 z_i) + (u_{0i} + u_{1i}x_{ij} + \varepsilon_{ij})$

3. Write down the equation of the population-averaged (marginal) model and the subject-specific (conditional) model.

Marginal model: $Y_{ij} = (\beta_0 + \beta_1 x_{ij} + \beta_2 z_i)$
Conditional model: $Y_{ij} = (\beta_0 + \beta_1 x_{ij} + \beta_2 z_i) + (u_{0i} + u_{1i}x_{ij})$

4. Repeat the last step but this time separately write the marginal and conditional models as two equations, one for planktotrophic larvae and the other for lecithotrophic larvae.

```
contrasts(sp$feeding.type)
  P
L 0   #for lecithotrophic larvae, z_i = 0.
P 1   #for planktotrophic larvae, z_i = 1.
```

Marginal model, P: $Y_{ij} = (\beta_0 + \beta_1 x_{ij} + \beta_2 z_i)$
Marginal model, L: $Y_{ij} = (\beta_0 + \beta_1 x_{ij})$
Conditional model, P: $Y_{ij} = (\beta_0 + \beta_1 x_{ij} + \beta_2 z_i) + (u_{0i} + u_{1i}x_{ij})$
Conditional model, L: $Y_{ij} = (\beta_0 + \beta_1 x_{ij}) + (u_{0i} + u_{1i}x_{ij})$

Question 4

Calculate an appropriate pseudo-R^2 to quantify the importance of the variable **feeding.type** for the model you chose in Question 2.

```
VarCorr(model3) #model chosen in Question 2, with feeding.type as a predictor
species = pdLogChol(I(lntemp - log(15)))
                      Variance   StdDev    Corr
(Intercept)          0.75250588 0.8674710 (Intr)
I(lntemp - log(15)) 0.28143057 0.5305003 -0.479
Residual             0.02303040 0.1517577
VarCorr(model2.5)  #model without feeding.type as a predictor
species = pdLogChol(I(lntemp - log(15)))
                      Variance   StdDev    Corr
(Intercept)          0.85672248 0.9255930 (Intr)
I(lntemp - log(15)) 0.28109558 0.5301845 -0.49
Residual             0.02310588 0.1520062
```

```
#doing the pseudo-R2
(0.85672248-0.75250588)/ 0.85672248
[1] 0.1216457 #a 12% difference in variability.
```

By allowing feeding.type to modify the intercept, the value for the intercept variance has decreased from 0.857 to 0.753, meaning that the new model (model3) does a better job of explaining the variance than the previous model (model2). The pseudo-R2 gives us a measure of how much more variance the new model explains that the old model did not, which is 0.122. At first this doesn't seem like a considerable improvement, but it is, considering that the data already has a wide range of scatter (see Question 5).

Question 5

Plot the equation of the population-averaged model over the range of the data using the model you chose as being best in Question 2. Clearly indicate the two larval feeding types on your graph.

```
contrasts(sp$feeding.type)
  P
L 0    #for lecithotrophic larvae, z_i = 0.
P 1    #for planktotrophic larvae, z_i = 1.

fixef(model3)
      (Intercept) I(lntemp - log(15))        feeding.typeP
        2.6374882              -1.4489601             0.6773869
```

Population-averaged model, P: $Y_{ij} = (\beta_0 + \beta_1 x_{ij} + \beta_2 z_i)$
```
P.func<-function(x) (fixef(model3)[1]+fixef(model3)[3]) +
fixef(model3)[2]*(x-log(15))
```

Population-averaged model, L: $Y_{ij} = (\beta_0 + \beta_1 x_{ij})$
```
L.func<-function(x) fixef(model3)[1] + fixef(model3)[2]*(x-log(15))
```

Plotting the lines for both feeding types, showing the x and y intercepts

```
plot(sp$lntemp,sp$lnPLD, xlab='lntemp', ylab='lnPLD', xlim=c(0,5.5),
ylim=c(0,8))

lines(log(seq(0.5,200,10)), P.func(log(seq(0.5,200,10))), col=2)

lines(log(seq(0.5,200,10)), L.func(log(seq(0.5,200,10))), col=3)

legend(3,7,c('Planktotrophic','Lecithotrophic'), lty=c(1,1), col=c(2,3),
bty='n', cex=c(.8,.8))

mtext('Marginal model: feeding type affects the intercept', side=3, line=.5)
```

Marginal model: feeding type affects the intercept
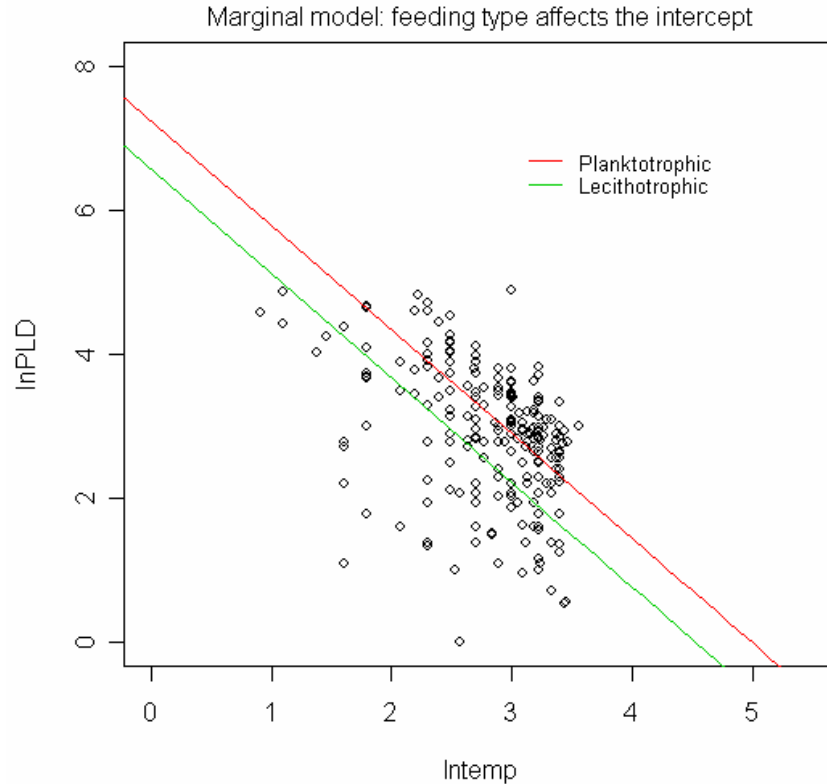
In words, what does your model say about the differences in the relationship between temperature and pelagic larval duration time for the two feeding types? Be as specific as you can be.

Because the feeding type is a categorical variable with two values (P and L), the model produces two lines. These lines are actually averages for 74 different lines (one for each species), grouped by feeding type. They suggest that at any temperature, the pelagic larval duration time for planktotrophic larvae is higher than the PLD for lecithotrophic larvae by a value equal to the exponentiated coefficient for feeding type, e^0.6773869. Also, the model suggests both feeding types have the same negative relationship between temperature and PLD—with each unit that the temperature increases, the PLD decreases by e^-1.4489601.

The filter-feeding planktotrophic larvae might have a higher PLD than the lecithotrophic ones because they might spend more time feeding in the water column as larvae, while lecithotrophic larvae have egg yolks for food, and don't need to remain in the water column. Also, maybe the yolk makes them heavier so they sink faster.

Interestingly, there don't seem to be any data points past `exp(max(sp$lntemp))` = 35 degrees on the graph, though there are lots of points clustered around that value of x. This might represent a temperature limit around 35 degrees. If so, there shouldn't be a steady decrease in PLD all the way to e^5 (148 degrees), where the model suggests pelagic larval duration time should be zero. Past 35 or so degrees, increasing the temperature would make the larvae die instead of decreasing their PLD further. The model might have a good fit to the data, but it likely doesn't perform as well outside the range of the data it was fitted for.